

Reconnaissance de gestes dynamiques de la main

Bogdan IONESCU, Didier COQUIN, Patrick LAMBERT

Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance
Domaine Universitaire, BP 806, 74016 Annecy Cedex, France

Didier.Coquin@univ-savoie.fr, Patrick.Lambert@univ-savoie.fr,
Bogdan.Ionescu@univ-savoie.fr

Résumé – Nous proposons une nouvelle méthode de reconnaissance d'un geste dynamique de la main. Un geste dynamique est caractérisé par les signatures statiques de début et fin de geste (histogramme des orientations du gradient) et par la signature dynamique (superposition des squelettes de chacune des images composant la séquence). La reconnaissance du geste se fait par une mesure de similitude entre ces signatures et des signatures d'un alphabet connu.

Abstract – In this paper, a new method aiming at a dynamic hand gesture recognition is introduced. A dynamic hand gesture is characterized by static signatures of the beginning and the end of gesture (gradient orientation histograms) and by a dynamic signature (superposition of all the hand region skeletons). The recognition system is based on a measure of similitude between signatures.

1 Introduction

Ce travail se situe dans le contexte plus général du contrôle d'un poste de travail par reconnaissance des gestes de la main d'un opérateur. Des mesures primaires des angles des articulations des doigts de la main sont effectuées à l'aide d'un gant numérique. Au même niveau, un second module de traitement basé sur un système de reconnaissance par vision vient enrichir la prise de décision, permettant en particulier la prise en compte du mouvement global de la main. Dans ce papier nous nous intéressons à ce second module dont l'objectif est la reconnaissance d'un geste dynamique de la main. L'acquisition des séquences du geste est réalisée dans un environnement presque idéal puisque la main et l'avant-bras se détache facilement d'un fond homogène et sombre (voir Fig. 1). Dans un environnement moins favorable il faudrait mettre en oeuvre une méthode de segmentation basée sur la détection de la peau et une méthode de suivi de geste. Ce papier se focalise avant tout sur la méthode de reconnaissance.

L'analyse des gestes suppose que l'on fasse de l'analyse spatiale et de l'analyse temporelle. De nombreuses méthodes de reconnaissance de gestes existent dans la littérature [1, 2, 3]. Elles font le plus souvent apparaître une étape d'extraction et une étape de classification s'appuyant généralement sur un apprentissage effectué autour d'un alphabet réduit de gestes. Nous avons retenu deux approches: une première basée sur l'histogramme des orientations du gradient [4] et une seconde, originale, utilisant la superposition des squelettes de la main et de l'avant-bras, calculés sur l'ensemble de la séquence. Dans notre application, un geste dynamique est caractérisé par une séquence comportant de 20 à 50 images (voir Fig. 1 et Fig. 2).

L'histogramme des orientations du gradient est utilisé comme une signature statique calculée sur la première et la dernière image de la séquence. Le principe de calcul et les résultats sur certaines séquences sont exposés dans la section 2. Ces signatures statiques permettent de délimiter le geste dynamique. Le squelette est utilisé comme une signature dynamique cal-

culée sur chaque image de la séquence. La section 3 aborde la méthode d'extraction du squelette, et le calcul de la signature dynamique. C'est cette signature qui assure la reconnaissance du geste. Le principe de sa reconnaissance est décrit dans la section 4.

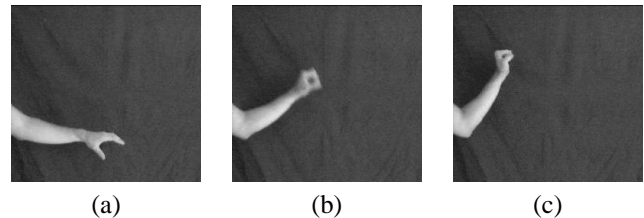


FIG. 1: (a) Début du geste, (b) image intermédiaire, (c) fin du geste.

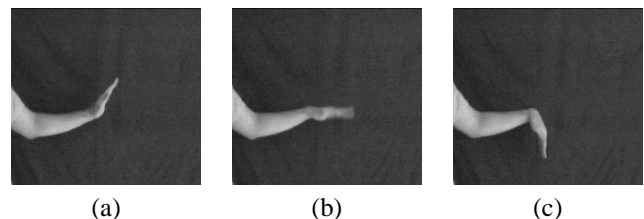


FIG. 2: (a) Début du geste, (b) image intermédiaire, (c) fin du geste.

2 Signatures statiques

Pour chaque geste on utilise comme signature statique, l'histogramme des orientations du gradient. Comme le fond n'est pas forcément homogène, et pour ne pas perturber les valeurs de l'histogramme, le calcul est réalisé uniquement sur les pixels de l'avant-bras et de la main. Cette première phase comprend 4 étapes :

- élimination automatique du fond par une méthode de binarisation automatique suivi d'un enchaînement de quelques opéra-

teurs morphologiques (ouverture, fermeture, remplissage des trous, élimination de petites régions),

- calcul du gradient sur chaque pixel de la main et de l'avant bras

- élimination des pixels dont le module du gradient est inférieur à K fois le module moyen. Des tests ont conduit à prendre $K = 1.2$,

- calcul de l'histogramme des orientations du gradient en utilisant un pas angulaire de 1 degré.

La figure (Fig. 3a, 3b) montre ces histogrammes obtenus sur les images de début et de fin de la séquence "GET".

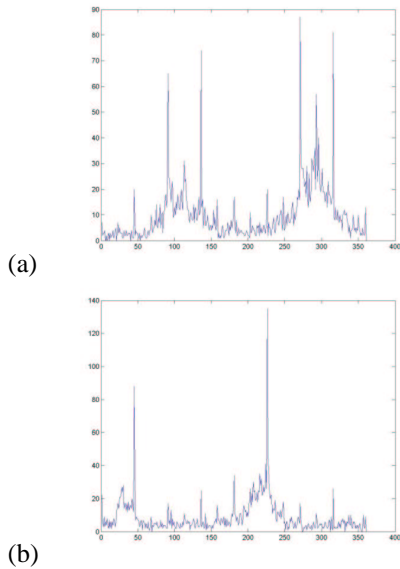


FIG. 3: Signature statique de la séquence "GET" (a - début du geste, b - fin du geste)

3 Signature dynamique

La signature dynamique est constituée de la superposition des squelettes de l'avant-bras et de la main extraits de chaque image binaire de la séquence du geste. Nous avons utilisé la méthode développée dans [5] pour extraire le squelette simplifié pour chaque image de la séquence. Nous rappelons ici les principales étapes de la méthode de squelettisation:

- calcul de l'image de distances en utilisant un masque de chanfrein adapté au maillage carré. Pour une question de rapidité nous avons utilisé un masque de taille 3x3 avec comme coefficients: 5 pour les directions verticale et horizontale, puis 7 pour la direction diagonale. Un algorithme séquentiel en 2 passages sur l'image est utilisé [6].

- extraction de l'axe médian et construction des chemins de connexion: par un parcours séquentiel les maxima locaux sont détectés par de simples tests dans un voisinage du pixel courant. Si nous sommes en présence d'un maximum local le parcours est interrompu temporairement et les chemins de connexion sont construits à partir de ce pixel.

- remplissage des faux trous: les faux trous sont détectés lors d'un parcours de l'image du squelette et sont remplis en vue de préserver la topologie de l'objet initial.

- élagage: les branches non significatives provenant de la discrétisation de l'objet sont supprimées par seuillage.

embellissement: les pixels formant des dents ou ne servant pas à la connexion sont supprimés. Les zig-zag sont alignés.

Ces différentes étapes sont présentées dans la figure suivante:

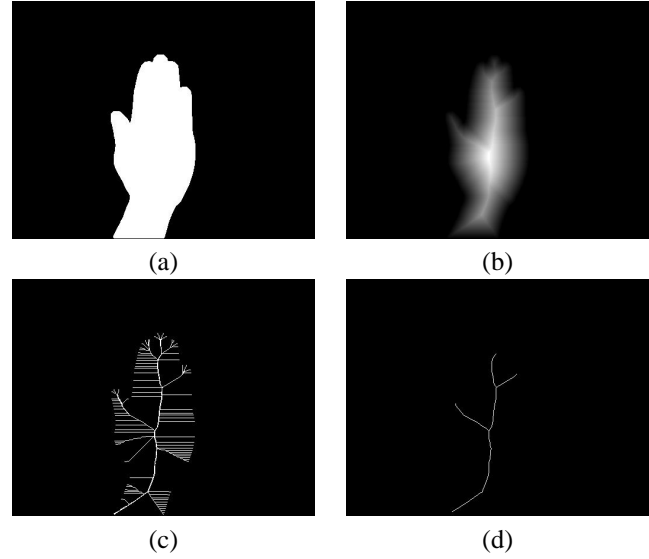


FIG. 4: (a) Image binaire, (b) Image de distances, (c) le squelette, (d) squelette simplifié.

La figure 5 montre les signatures dynamiques des gestes "GET" et "SLOW".

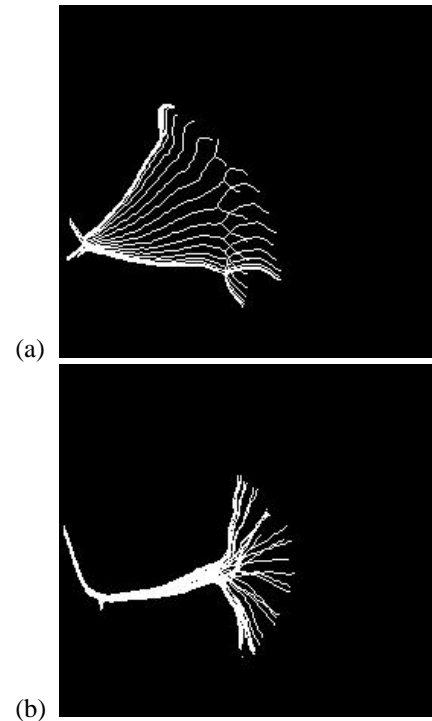


FIG. 5: Signature dynamique de la séquence "GET" (a) et "SLOW" (b)

Il faut noter que cette signature dynamique originale présente l'avantage de résumer l'ensemble du geste en une seule image. En revanche, elle ne permet pas de traduire l'information sur la chronologie du geste. Ceci n'est pas gênant pour des gestes

simples comme c'est le cas dans notre application.

4 Reconnaissance d'un geste dynamique

La reconnaissance du geste s'effectue en 2 temps. D'abord, on cherche à reconnaître les signatures statiques de début et de fin de geste. Ces signatures, une fois reconnues, permettent de délimiter le calcul de la signature dynamique. Dans un deuxième temps, on procède à la reconnaissance de cette signature dynamique

4.1 Principe

Le principe de la reconnaissance d'un geste repose sur la comparaison entre les signatures du geste courant et les signatures des séquences d'apprentissage correspondant à un alphabet de gestes connus. Le geste reconnu est associé à la distance la plus petite.

Pour la signature statique, nous utilisons simplement la distance Euclidienne entre les histogrammes des orientations du gradient.

Pour la signature dynamique, nous devons comparer deux images binaires représentant la superposition des squelettes de la séquence courante et des séquences d'apprentissage. La comparaison se fait en calculant la distance de Baddeley [7] [8].

Rappelons que la distance de Baddeley entre deux images binaires A et B est définie par:

$$D(A, B) = \left[\frac{1}{\text{card}(S)} \sum_{s \in S} |d_A(s) - d_B(s)|^E \right]^{\frac{1}{E}} \quad (1)$$

où $s = (x, y)$ est un pixel du support S , E est un réel supérieur à 1 (on prend généralement $E = 2$), $d_A(s)$ est la distance entre le pixel s et l'ensemble de référence (ensemble binaire caractérisant l'image A). Cette distance sera approximée en calculant une transformation de distance avec un masque de chanfrein de taille 5x5 [6].

Le processus de reconnaissance des gestes dynamiques de la main est illustré en figure 6.

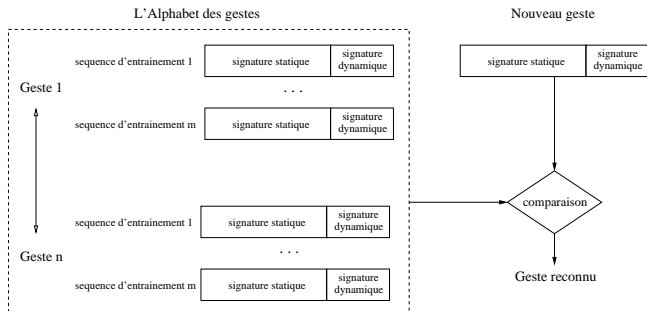


FIG. 6: Processus de reconnaissance.

4.2 Résultats

Cette méthode a été testée sur 40 séquences de geste, pour un alphabet de 10 gestes. Les performances obtenues dépendent de la nature du geste (dynamique ou statique) et du nombre de séquences d'apprentissage utilisées pour un même geste.

En effet, pour rendre la méthode robuste à la variabilité pouvant apparaître sur un même geste, la base d'apprentissage peut éventuellement comporter plusieurs modèles pour un même geste.

4.2.1 signature statique

Le tableau ci-dessous (Tab. 1) donne, pour les signatures statiques de début et de fin de geste, l'évolution du taux de reconnaissance en fonction du nombre de modèles utilisés pour un même geste.

TAB. 1: Evolution du taux de reconnaissance des signatures statiques (en %) en fonction du nombre de séquences d'apprentissage.

Nb séq. apprentissage	geste statique de début	geste statique de fin
1	88.4	96.1
2	77.7	91.6
3	91.6	94.4
4	91.6	94.4
5	97.2	94.4
6	100	94.4

On constate, sur les expériences faites, que le taux de reconnaissance est élevé, mais qu'il n'atteint pas les 100%. Par contre, l'utilisation de plusieurs modèles pour un même geste n'améliore pas sensiblement les performances.

Le traitement de ces signatures est rapide et permet d'envisager une application temps réel.

Il est important de remarquer que les erreurs obtenues ne correspondent pas à des non-détections de geste de début ou de fin, mais à des confusions entre gestes de début ou entre gestes de fin. La reconnaissance de ces signatures statiques ne servant qu'à délimiter la signature dynamique, ces erreurs sont donc sans conséquence sur la détection du geste dynamique.

4.2.2 signature dynamique

Le tableau ci-dessous (Tab. 2) donne, pour la signature dynamique, l'évolution du taux de reconnaissance en fonction du nombre de modèles utilisés pour un même geste.

TAB. 2: Evolution du taux de reconnaissance de la signature dynamique (en %) en fonction du nombre de séquences d'apprentissage.

Nb séq. apprentissage	geste dynamique
1	97.2
2	100
3	100
4	100
5	100
6	100

On constate, sur les expériences faites, que le taux de reconnaissance est égal à 100% dès que l'on utilise au moins deux modèles d'apprentissage pour un même geste.

La signature dynamique aboutit à de meilleures performances que les signatures statiques, ce qui semble naturel puisqu'elle

utilise plusieurs images de la séquence. En revanche le traitement de ces signatures est plus long et ne s'effectue pas, pour le moment, en temps réel.

5 Conclusion et perspectives

Nous avons proposé une méthode de reconnaissance de gestes de la main basée sur l'utilisation conjointe de signatures statiques et d'une signature dynamique. Les signatures statiques, composées de l'histogramme des orientations du gradient de l'image de ma main et de l'avant bras, permettent de détecter la partie dynamique utile. La signature dynamique proposée est une solution originale qui consiste à superposer les squelettes des différentes images composant le geste.

La méthode proposée a été testée sur une quarantaine de gestes pris parmi un alphabet de 10 gestes simples. Moyennant l'utilisation de plusieurs séquences d'apprentissage pour un même geste, le taux de reconnaissance obtenu atteint les 100%.

Les perspectives de ce travail sont doubles:

- d'une part, elles concernent l'amélioration des performances en terme de temps de calcul. Si la reconnaissance des gestes statiques est rapide et permet d'envisager une implantation temps réel, la reconnaissance du geste dynamique nécessite une phase d'optimisation et d'accélération des traitements.
- d'autre part, les résultats fournis par l'analyse des images doivent être fusionnés avec ceux fournis par le gant numérique.

Références

- [1] V. Pavlovic, R. Sharma, T. Huang. *Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review*. IEEE PAMI, vol. 19 (7), 1997.
- [2] L. Bretzner, I. Laptev, T. Lindeberg. *Hand Gesture Recognition using Multi-Scale Colour Feature, Hierarchical Models and Particle filtering*. In Proc. International Conference on Automatic Face and Gesture Recognition, June 2002.
- [3] Byung-Woo Min, Ho-Sub Yoon, Jung Soh, Takeshi Ohashi, Toshiaki Ejima. *Visual Recognition of Static/Dynamic Gesture: Gesture-Driven Editing System*. Journal of Visual Languages and Computing 10(3): 291-309 (1999)
- [4] William T. Freeman and Michal Roth, *Orientation Histograms for Hand Gesture Recognition*, Mitsubishi Electric Research Laboratories, Cambridge Research centre, Technical Report -94-03a, Dec. 1995.
- [5] Y. Chehadah, D. Coquin, Ph. Bolon. *A skeletonization algorithm using chamfer distance transformation adapted to rectangular grid*. IEEE International Conference on Pattern Recognition, Vol. 2, pp. 131-135, August 1996.
- [6] G. Borgefors. *Distance Transformations in arbitrary dimensions*. CVGIP, vol. 27, pp. 321-345, 1984.
- [7] A.J. Baddeley. *An error metric for binary images*. Robust Computer Vision, Wichmann, Karlsruhe, pp. 59-78, 1992.
- [8] D. Coquin, Ph. Bolon. *Applications of Baddeley's distance to dissimilarity measurement between gray scale images*. Pattern Recognition Letters, Vol. 22, pp. 1483-1502, 2001.

6 Annexe

Présentation de 8 gestes utilisés dans l'alphabet, auquel il faut ajouter les gestes Get et Slow voir Fig. 1 et Fig. 2.

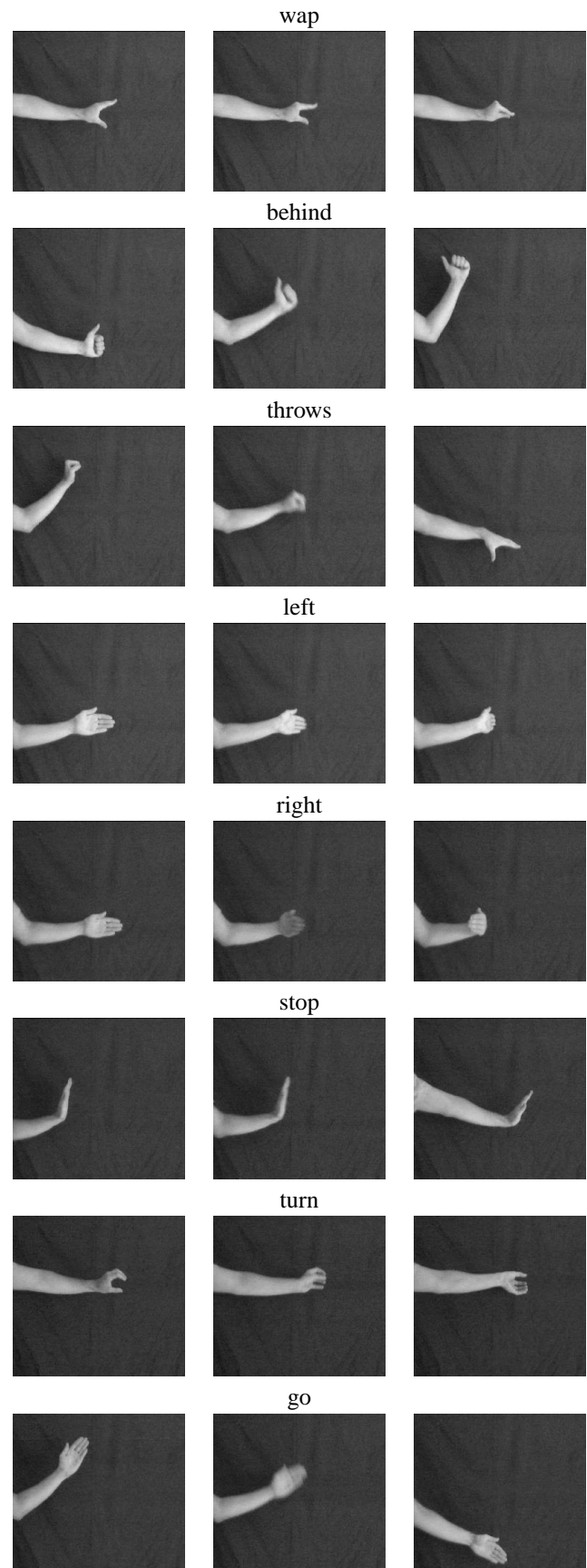


FIG. 7: Alphabet des gestes dynamiques.